International Journal of Inventive Engineering and Sciences (IJIES)
ISSN: 2319-9598 (Online), Volume-11 Issue-5, May 2024

DNN-PolSAR: Urban Image Segmentation and Classification using Polarimetric SAR based on DNNs

Soumyadip Sarkar, Farhan Hai Khan, Shobhit Kumar, Tamesh Halder, Dipjyoti Paul, Debashish Chakravarty

Abstract. ***Synthetic Aperture Radar (SAR) image segmentation and classification are popular techniques for learning and detecting objects such as buildings, trees, monuments, crops, water bodies, hills, etc. The SAR technique is being utilised for urban development and city planning, building control of municipal objects, identifying optimal locations, and detecting changes in existing systems, among other applications, by leveraging polarimetry based on Deep Neural Networks. In this paper, we propose a technique for urban image segmentation and Classification using Polarimetric SAR based on Deep Neural Networks (DNN-PolSAR). In our proposed DNN-PolSAR technique, we utilise Mask-RCNN, LinkNet, FPN, and PSP-Net as model architectures, while ResNet-50, ResNet-101, ResNet-152, and VGG-19 are employed as backbone networks.We first apply polarimetric decomposition to airborne Uninhabited Aerial Vehicle Synthetic Aperture (UAVSAR) images of urban areas, and then the decomposed images are fed to DNNs for segmentation and classification. We then simulate DNN-PolSAR considering different hyperparameters and compare the obtained scores of these hyperparameters against the used model architectures and backbone networks. In comparison, it is found that DNN-PolSAR, based on the FPN model with ResNet152, performed the best for segmentation and classification. The mean Average Precision (mAP) score of the DNN-PolSAR based on FPN with a pixel accuracy of 90.9% is 0.823, which outperforms other Deep Learning models.***

Keywords: Polarimetric SAR, FPN, PSPNet, Mask-RCNN, LinkNet, Image Segmentation.

I. INTRODUCTION

Segmentation and classification of an image is a process of dividing and categorising the image into distinct parts based on predefined categories of objects. In this process, each pixel in an image is categorized based on the predefined labels of objects.

Manuscript received on 12 April 2024 | Revised Manuscript received on 11 May 2024 | Manuscript Accepted on 15 May 2024 | Manuscript published on 30 May 2024.
**Correspondence Author(s)*
Soumyadip Sarkar, Institute of Engineering & Management, Kolkata, India. Email: soumya997.sarkar@gmail.com
Farhan Hai Khan, Institute of Engineering & Management, Kolkata, India, Email: njrfarhandasilva10@gmail.com
Shobhit Kumar, Institute of Engineering & Management, Kolkata, India, Email: shobhit.course@gmail.com
Tamesh Halder*, Department of Mining Engineering, Indian Institute of Technology, Kharagpur, India. E-mail: haldertamesh@iitkgp.ac.in, ORCID ID: 0000-0002-9363-1471
Dipjyoti Paul, Department of Computer Science, University of Crete, Heraklion, Crete, Greece. Email: dipjyoti92@gmail.com
Debashish Chakravarty, Department of Mining Engineering, Indian Institute of Technology, Kharagpur, India, Email: profdciitkgp@gmail.com

Image segmentation has historicallybeen used primarily for recognising scenes in which similar objects can be accurately placed. However, image segmentation is successfully applied in various fields, including medical imaging and autonomous driving. Therefore, image segmentation can also be used for satellite images and Polarimetric SAR (PolSAR) of urban cover areas for categorization and analysis [1],[2].

PolSAR is a widely used technique in remote sensing, employed in various applications, including segregation and classification in GIS. It is also used for mapping areas such as forests, vegetation, and urbanised regions. Data generated from PolSAR provides SAR resolutions, which help to understand images in terms of scattering components, including surface scattering, volume scattering, helix scattering, double-bounce scattering, and wire scattering. Based on these scattering components, PolSAR facilitates the classification of objects. For example, it is seen that PolSAR generates more prominent helix and double-bounce scattering components for images of urban areas [3]–[5]. In this work, we consider double-bounce scattering and helix scattering components for classifying objects in images of urban areas. However, with the growth of urbanisation and an increasing population in urban areas, tracking, studying, and analysing urban cover areas have become essential, particularly in terms of locating and classifying objects such as buildings, crops, water bodies, and hills. Therefore, accurately locating and classifying different objects using images of urban areas is crucial for designing efficient and reliable solutions [6]. However, urban image segmentation and classification are very challenging tasks, even when using SAR polarimetry. This is because urban cover relatively shows a small part of the total surface. Fortunately, a vast collection of freely available satellite imagery datasets is available, which can be used for image segmentation and classification of urban cover areas.

Image segmentation and classification of urban cover areas using PolSAR is a challenging task due to urban structures, whose orientation is not in line of sight (LoS) of the radar. However, recognition of such areas is essential for several reasons, including disaster relief, urban planning, and environmental monitoring. But, it is not possible to feed the scattering of images of urban cover areas taken using the Uninhabited Aerial Vehicle Synthetic Aperture Radar (UAVSAR) into a neural network.

This is because it is necessary to employ a set of image decompositions to retrieve various information using scattering, such as surface scattering, double-bounce scattering, volume scattering, helix scattering, and wire scattering, among others. It is also necessary to identify different areas, such as grassland, urban areas, hills, etc., with distinct scattering information and components from the images, so that the data obtained from scattering becomes significant. A scatteringcomponent enables us to determine the type of area captured in particular photos. A grassland,for example, may have high values for surface scattering, while an urban area may have high valuesfor both double bounce and helix scattering. By applying these decomposition techniques to the UAVSAR raw scattering matrix elements, different areas tend to exhibit distinct characteristics, which can be utilised for image segmentation and classification. Based on the principle mentioned above, in this paper, we present a technique for Polarimetric SAR (PolSAR) image segmentation and classification of Radar Satellite Imagery of urban areas using Deep Neural Networks (DNNs) such as PSPNet, LinkNet, FPN, and Mask-RCNN based on different backbone networks, such as Efficient-Net, DenseNet, MobileNet, Inception, ResNet, and VGG19 (discussed in Subsection 3.1). Inour proposed technique, we first apply polarimetric decomposition to airborne UAVSAR images of urban areas, and then the decomposed images are fed to DNNs for segmentation and classification. We simulate our proposed technique and obtain simulation results using different DNNs accordingly. The significant contributions of this work are as follows:

We propose a technique that combines models and backbone networks for urban classification and perform a rigorous evaluation of all machine learning classifiers in the field of Remote Sensing. We propose a technique to identify and detect buildings, grassland, and hills from the PolSAR images.

- Presented and described the best and most effective Deep Learning methods for PolSAR image segmentation and classification.
- We obtain simulation results for our proposed technique using different backbone networks, including EfficientNet, DenseNet, MobileNet, Inception, ResNet, and VGG19.
- We conducted an extensive comparison and discussion of the current state-of-the-art models on the same datasets for the segmentation and classification of urban area covers.

The remainder of the paper is organised as follows. Section 2 presents Related Work in the area of satellite image segmentation and classification. In Section 3, we discuss the architectures and backbones of the models used in our paper. An overview of the datasets used for training and validation is discussed in Section 4. In Section 5, the results of experimentation based on different databases, along with a discussion of the findings, are presented. Finally, we concluded our paper in Section6.

II. RELATED WORKS

Segmentation of PolSAR images of urban cover areas presents unique challenges, including partial surface visibility and diverse scattering mechanisms. However, the only good thing is that the area structure of these images is well-defined. However, the design and development of algorithms for analysing and classifying images require focused research to open up possibilities for the application of Remote Sensing in various fields. Given this, Several works have been proposed for image analysis and classification using semantic segmentation. In this section, we briefly present and discuss some of the advancements in classification approaches for PolSAR, as well as some examples where architectures from a different area of study have been successfullyapplied in remote sensing. A recent study conducted by De et al. [7] aimed to develop a deep learning-based novel technique for classifying urban areas. The information in the augmented dataset used in this work is transformed using a stacked autoencoder before being fed to a neural network for classification. Thistechnique achieved an accuracy of 91.3%, representing a performance improvement over existing techniques at the time. In [8], Cui et al. proposed an architecture comprising a Dense Attention Pyramid Network (DAPN), a Region Proposal Network (RPN), and a detection network for multi-scale ship detection in SAR images. Here, DAPN was used to extract multi-scale fused features for generating and detecting, which are then used in the subsequent iterations of the technique. Top-down, densely connected networks are used to obtain concatenated feature maps from lower layers. The proposed method provided an accuracy of 89.8%, which was 11% higher than the previous models on the SAR ship detection data set (SSSD). DAPN was also 20% faster than the faster R-CNN [9]. They also demonstrated that the top-down pyramid structure with attention is highly effective in obtaining feature maps that contain more spatial and semantic information. Recently, Mohanty et al. presented applications of Mask-RCNN [10] for segmenting and detecting buildings in Google Maps Satellite Imagery Data. The authors found the results to be impressive, with a final loss value of 0.15 for the instance image segmentation model. Wang et al. explored the problems in classifying PolSAR images due to the presence of nonlinear data.11 This study proposes a kernel sparse representation-based classification approach. This kernel function technique solves the problems caused by nonlinear features. This helps achieve more accurate results in the classification task. This study used an Airborne SAR dataset from San Francisco, United States of America. In [12], Femin et al. proposed an approach for detecting buildings using a CNN from satellite images. In this work, different building footprints were identified in images using a CNN method. The proposed work also detected different shapes and colours.

The detection accuracy by this approach for building was found to be 83%. On the other hand, Wang et al. introduced a deep feature extraction approach in [13], where a multilevel polarimetric feature vector is extracted using a PAO PTD CNN. The authors extracted superpixels usingsimple linear iterative clustering (SLIC) from the feature vector for the classification map. Finally, the result is obtained by combining the superpixel map and the deep feature classification vector, with a Kappa Score of 0.86. The authors of [14] noted that semantic segmentation can also be applied to high-resolution images.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

PolSAR images are processed using neural network architectures, such as MP-ResNet, which contains three concurrent semantic embedding branches and employs a multi-scale feature fusion design in the decoder to utilise each encoding branch.

The authors observed that MP-ResNet enhances the aggregation of contextual information compared to the baseline Fully Convolutional Network (FCN). The proposed method, based on MP-ResNet, surpasses numerous state-of-the-art methods in terms of accuracy, achieving a mean F1 score of 92.25% and an IoU of 89.60% in classification using the Gaofen Dataset. Zhao et. al. showed in [15] that segmentation can also be achieved using edge information based on spectral graphpartitioning. Here, the authors defined segmentation as a three-part process, namely edge information extraction, edge-based similarity matrix analysis, and normalised cut. This method overcame the pepper-salt phenomenon, along with much more complete boundaries of the segments.The method by Ouahabi et al. [16] aimed to improve segmentation efficiency without compromising accuracy using a Fully Convolutional dense Dilated Network model. Here, the authors found that low resolution and contrast, shadow interference, as well as differences in size and position of the abnormal tissue, are the challenges that hinder the process of segmenting ultrasound images. Yuanyuan et al. in their work17 explore how different classification algorithms are affected by the choice of polarimetric parameters such as Alpha, HAAlpha T11, Shannon entropy, VanZyl3 Vol, Neuman delta mod, Barnes2 T33, Barnes1 T33, and entropy.

III. BACKBONE NETWORK AND MODEL ARCHITECTURE

A. Backbone Networks

A backbone network is primarily used to extract network features for object classification and detection. In this paper, we utilise ResNet152 [18], ResNet101 [19], ResNet50 [20], and VGG-19 [20] as backbone networks for feature extraction from images.

B. Model Architectures

In this Subsection, we explain the model architectures used for classification in our work, including M-RCNN, PSPNet, FPN, and LinkNet. The model architectures classify the extracted features using the base model from the deep neural network backbone discussed in Section 3.1.

C. MR-CNN

The M-RCNN [21] was developed as an extension to the Faster-RCNN [9], which has been widely usedso far for various object detection purposes. The F-RCNN/M-RCNN as output yields an object's label along with the object's bounding box. F-RCNN uses a feature extractor block that extracts the features from the image. These features are then used to train the bounding box regressor andthe classifier. The M-RCNN, as the name suggests, extends F-RCNN by training a binary mask in parallel with the bounding box regressor and object classifier. The first stage of the Mask-RCNN (like theF-RCNN) is the Region Proposal Network (RPN). Each bounding box is paired with an objectnessscore, indicating the likelihood of the object being present. The second stage of the M-RCNN is referred to as the network's head. In F-RCNN, this head typically consists of a stack of convolutional layers anda dense layer for bounding box regression. M-RCNN, in parallel with this bounding box learning algorithm, uses a stack of convolutional layers for Mask representation. This parallel task makes it theoretically faster and more accurate than other segmentation models.

D. Feature Pyramid Network (FPN)

A FPN is a fully convolutional feature extractor that takes a single-scale image of any size as input and produces correspondingly sized feature maps at several layers. [22] The model comprises two distinctive parts: a conventional convolutional network (such as VGG19 or ResNet50), which acts as a feature extractor, and a deconvolutional network with compatible feature sizes. However, there is a crucial difference between these two parts: the convolutional network operates from bottom to top, whereas the flow in the deconvolutional network operates from top to bottom. The blocks in the convolutional network are connected in the deconvolutional network by linear multiplication.The output of blocks in the deconvolutional layer is connected to individual convolution layers, which are not directly connected. These layers are transformed into a stack. This datasetundergoes upsampling and activation to produce an image map.

E. LinkNet

The LinkNet is a lightweight network architecture designed for performing segmentation tasks with a special focus on processing time [23]. Instead of a typical auto-encoder style segmentation model where the spatial semantics are first extracted using encoder blocks and then the decoder uses thisspatial information for spatial categorization. This method has a particular downside in terms of both computation and accuracy. The pooling and strided convolution used in encoders may result in some loss of spatial information. Instead, the LinkNet algorithm utilises skip connections from one encoder block to the corresponding block, thereby preventing information loss at each stage. This concept of semantic information preservation is similar to a U-Net, except that in this case, the encoder's results are combined with the corresponding decoder block's results, rather than performing feature concatenation. For experimentation, we will be using the model proposed in the original LinkNet paper. The model uses four encoder blocks and four corresponding decoder blocks. There are two special blocks of fully convolutional neural networks at the beginning and end of the network to preserve the image's dimensions.

a. Pyramid Scene Parsing Network (PSPNet)

The PSPNet of [24] is a model used for semantic segmentation. Its speciality is that it uses a pyramidparsing module. This module utilises region-based context aggregation to leverage global context information.

The final predictions are made more reliable due to the presence of both local and global clues. Given an input image, the feature map can be extracted using a pre-trained CNN with a dilated network strategy.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

The final size of the feature map is reduced to 1/8thof the input image. A pyramid pooling module is then applied on top of the map to gather context information. A four-level pyramid is used, where the pooling kernels cover the entire image, half of the image, and small parts of the image.

The results from the pooling kernels are then concatenated to form a global prior. In the next step, this prior is concatenated to the original feature map. The obtainedresult is finally passed through a stack of convolutional layers to generate the final prediction.

IV. DATASETS AND USAGE IN PROPOSED TECHNIQUE

Datasets play a crucial role in machine learning algorithms for segmentation and classification. In our proposed technique, too, datasets play a significant role in the segmentation of Classification of images of urban cover areas.

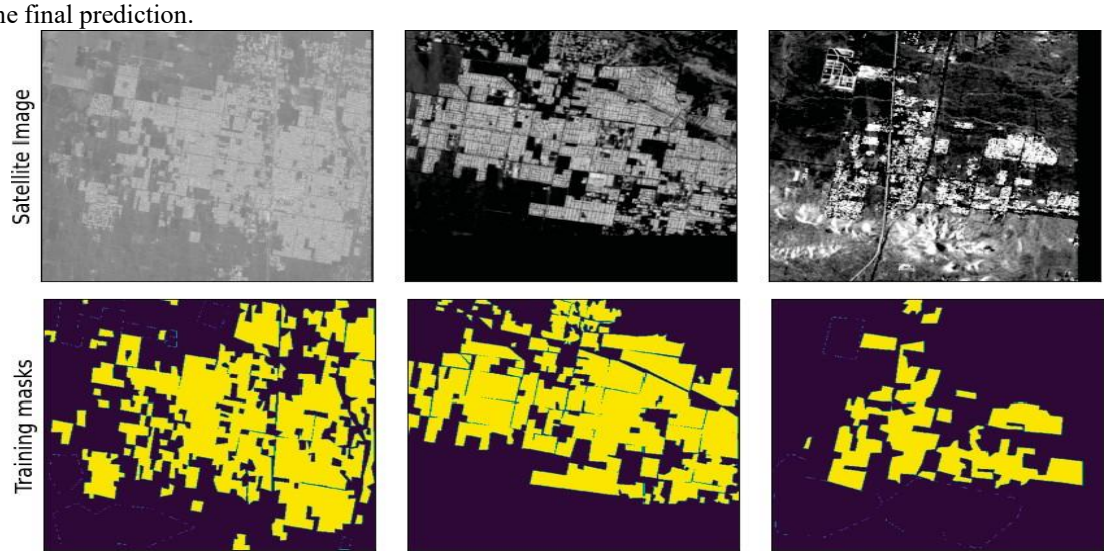


Figure 1: Sample Datasets and their Corresponding Masks for Urban Areas

We mentioned in Section 1 that a vast collection of satellite imagery datasets of urban cover areas is readily available for image segmentation and classification. Therefore, to train our proposed algorithm, we utilised PolSAR images of Lancaster, Palmdale, and Rosamond cities from airborne UAVSAR. However, we have considered only building classes for semantic and instance segmentation from these datasets usingDeep Learning over various polarimetric decompositions. It is also worth mentioning that, similar to [17], we have used various polarisation parameters, such as Alpha, HAAlpha T11, Shannon entropy, VanZyl3 Vol, Neuman delta mod, Barnes2 T33, Barnes1 T33, and entropy, to improve the classification accuracy in our proposed technique. We utilise the PolSARPro v6.0 Software Suite [40] for the decomposition results in our proposed work. In Table ??, we show all the decomposition meth- ods and corresponding polarimetric parameters those were applied on the datasets in our proposedtechnique. It is worth mentioning that we also performed image augmentation using random rotation and image flipping to generate more data before passing it through the model. We generated three transformed images from each image with a size of 1331 x 1101 to enhance our datasets. The improved datasets are used for training based on PolSAR images of Lancaster, Palmdale, and Rosamond cities in the USA. The reason for using datasets from different cities is to increase segmentation accuracy by introducing variance in the datasets. Details of the used datasets are given in Table 2. As shown in Table 2, there are 71, 60, and 50 training datasets for Lancaster, Palmdale, and Rosamond, respectively. But, we used 33, 64, and 17 test datasets for Lancaster, Palmdale, and Rosamond, respectively.

Table 1 Shows Sample Datasets of Satellite Images (The Upper Parts of the Figure) as well as

Decomposition Method	Polarimetric Parameter		
Cloude [25]	Cloud T11	Cloud T22	Cloud T33
H/A/Alpha [26]	Entropy H/A/A T11	Anisotropy H/A/A T22	Shannon Entropy H/A/A T33
VanZyl3 [26]	VanZyl3 Vol	VanZyl3 Odd	VanZyl3 Dbl
Neuman [27]	Neuman delta mod	Neuman delta pha	Neuman tau
FreeMan2 [28]	FreeMan2 Vol	FreeMan2 Ground	
FreeMan [29]	FreeMan Vol	FreeMan Odd	Freeman Dbl
Huyen [30]	Huyen T11	Huyen T22	Huyen T33
Bhattacharya [31]	Frey Dbl	Frey Hlx	Frey Odd
Singh [32]	Singh 6SD1	Singh G4U2 Vol	Singh G4U2 Odd
Barnes1 [33]	Barnes1 T11	Barnes1 T22	Barnes2 T33
Barnes2 [33]	Barnes2 T11	Barnes2 T22	Barnes2 T33
Pauli [25]	Pauli a	Pauli b	Pauli c
Holm1[34]	Holm1 T11	Holm1 T22	Holm1 T33
Holm2 [34]	Holm2 T11	Holm2 T22	Holm2 T33
Arri3 NNED [35]	Arii NNED Vol	Arii NNED Odd	Arii NNED Dbl
An Yang3 [36]	An Yang3 Vol	An Yang3 Odd	An Yang3 Dbl
An Yang4 [37]	An Yang4 Vol	An Yang4 Odd	An Yang4 Dbl
Yamaguchi3 [38]	Yamaguchi3 Vol	Yamaguchi3 Odd	Yamaguchi3 Dbl
Yamaguchi4 [39]	Yamaguchi3 Vol	Yamaguchi3 Odd	Yamaguchi3 Dbl

Retrieval Number: 100.1/ijies.E444813050624
DOI: 10.35940/ijies.E4448.11050524
Journal Website: www.ijies.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Table 2: Description of the Datasets

Location	Coordinates		Region/Country	Datasets	
	Latitude	Longitude		Train	Test
Lancaster	40.037° N	76.305° W	Pennsylvania, USA	71	33
Rosamond	34.8641° N	118.1634° W	Karen County, California, USA	60	64
Palmdale	34.3452° N	118.62° W	Los Angeles, California, USA	50	17

Corresponding masks (the lower parts of the Figure) of the satellite images. From Figure 1, it can be seen that our proposed technique, based on the datasets, correctly segments and classifies urban areas in the photos. In the lower part of Figure 1, the yellow-coloured masks represent the presence of the metropolitan regions in the given satellite images. Details of the results obtained with our proposed technique, along with a discussion on the results, are presented in Section 6.2.

A. Proposed Technique

We employ Deep Neural Networks (DNNs) such as PSPNet, LinkNet, FPN, and M-RCNN, based on various backbone networks including EfficientNet, DenseNet, MobileNet, Inception, ResNet, and VGG19, to segment and classify Radar Satellite Imagery of urban areas. The block diagram of theproposed technique is shown in Figure 2. In the proposed method, polarimetric decomposition and the refined Lee filter are applied to the airborne UAVSAR images of urban areas. The decomposed images are then fed to DNNs along with their respective backbone networks for segmentation and classification. From Figure 2, it can be seen that the FPN, PSP Net, and Link Net utilise three different backbone networks, namely ResNet50, ResNet152, and VGG19, whereas M-RCNN employs ResNet50, ResNet101, and VGG19 as its backbone networks.

B. Simulation Studies

In this Section, we present simulation results of our proposed technique and provide a discussion on these results. It is worth mentioning that the primary motivation of our work is to understand the learning capacity and rate of convergence of image segmentation and classification about the architectures above, using different backbone networks. To achieve this, we have considered four different model architectures, as well as four different backbone networks, to obtainunbiased results. We have used hyper-parameters such as *Intersection Over Union (IoU) score*, *Pixel Accuracy*, *F1 Score*, *Cohen's Kappa Score*, *Area Under the Curve*, *Recall*, *Precision*, and *Mean Average Precision(mAP)* as performance metrics to obtain simulation results of learning capacity and rate of convergence by by our proposed technique. We have also discussed these metrics to draw a performance comparison of different architectures and backbone networks.

C. Simulation Environment

We have simulated our proposed technique using Python 3 on a Kaggle Colab notebook and R language on a computer with 64 GB of RAM. Simulation is conducted using model architectures, namely M-RCNN, FPN, LinkNet, and PSPNet, against the ResNet-152, ResNet-101, ResNet-50, and VGG-19 backbones. However, we provided the results of the best-performing backbone networks for each of the considered model architectures.

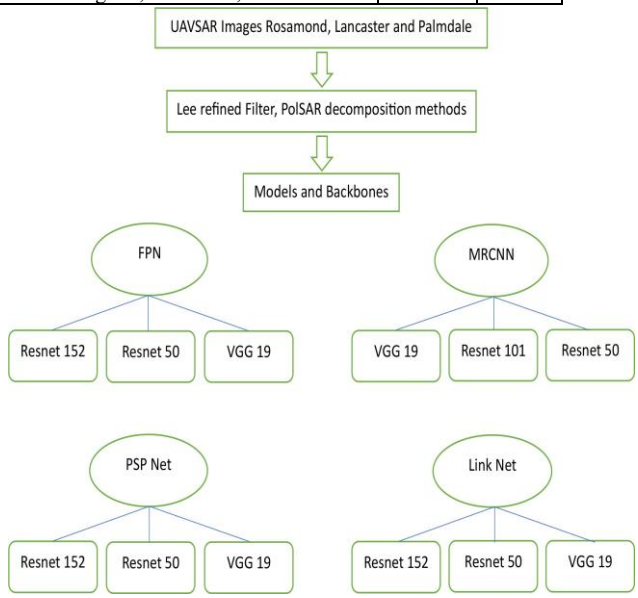


Figure 2: Block Diagram of Our Proposed Scheme

V. RESULTS AND DISCUSSION

In this section, we present simulation results and a discussion. Simulation results are presented considering popular hyperparameters for the deep learning algorithms used in our approach. Timely convergence of deep learning algorithms is crucial and essential. We also presented an analysis of the convergence of the DL algorithm used in our work. In the final subsection of the section, a case study is also given, considering the prediction accuracy of urban images.

A. Numerical Results

In Table 3, we present simulation results for Pixel Accuracy, Cohen's Kappa Score, IoU Score, mean Average Precision, and Area Under the Curve.

Recall, Precision, and F1 score hyperparameters for all considered model architectures and backbone networks are presented. From Table 3, it can be seen that Cohen's Kappa Score and IoU Score for FPN with ResNet-152 are the highest among all other scores. It can also be seen from Table 3 that mAP, AuC, and F1 Scores for FPN with ResNet152 and VCG19 are the highest, respectively. Therefore, it may be concluded that the FPN gives the best accuracy among all models proposed in this paper.

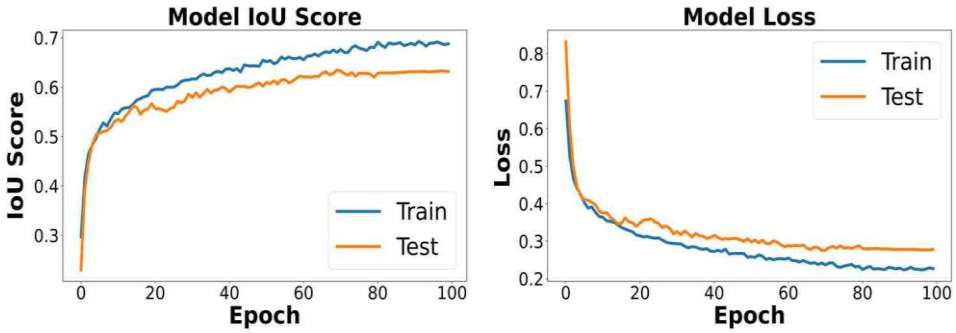
The high F1 score and AuC scores for the top three models confirm that the FPN architecture performsbest among all other architectures. It achieves pixel accuracy above 90% for three backbones:ResNet152, VGG-19, and ResNet50. On the other hand, it can be seen from Table 3 that the PixelAccuracy of 91% for LinkNet with ResNet152 is the highest. However, the values of Recall and Precision are highest for LinkNet with ResNet-50 and ResNet-152, respectively.

Table 3: Pixel Accuracy, Cohen's Kappa Score, IoU Score, mAP, AuC, Recall, Precision, and F1 score of all Considered Model Architectures and Backbone Networks

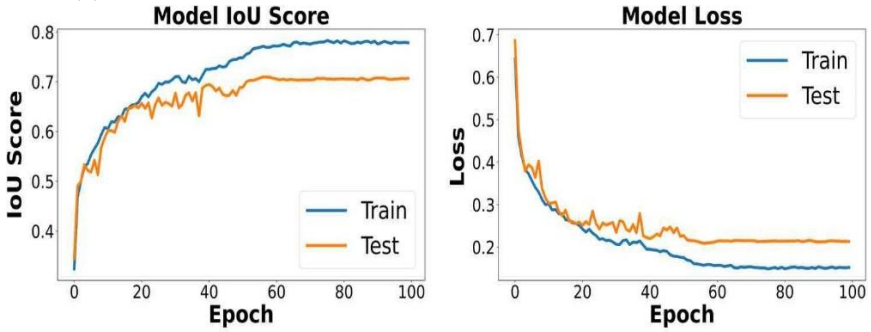
Model Architecture	Backbone Network	Pixel Accuracy	Cohen's Kappa Score	IoU Score	mAP	AuC	Recall	Precision	F1 Score
FPN	ResNet152	0.909	0.806	0.799	0.823	0.965	0.917	0.850	0.882
	ResNet50	0.901	0.786	0.774	0.808	0.963	0.879	0.861	0.870
	VGG-19	0.909	0.805	0.796	0.817	0.968	0.928	0.843	0.884
MR-CNN	ResNet101	0.897	0.781	0.780	0.809	0.950	0.897	0.839	0.867
	ResNet50	0.638	0.057	0.069	0.394	0.634	0.075	0.647	0.134
	VGG-19	0.811	0.621	0.667	0.672	0.937	0.973	0.671	0.794
PSPNet	ResNet152	0.885	0.755	0.753	0.768	0.960	0.934	0.795	0.859
	ResNet50	0.895	0.771	0.758	0.784	0.961	0.896	0.835	0.865
	VGG-19	0.893	0.772	0.763	0.788	0.957	0.907	0.824	0.864
LinkNet	ResNet152	0.910	0.805	0.791	0.822	0.966	0.895	0.868	0.881
	ResNet50	0.464	0.127	0.419	0.419	0.618	1.000	0.411	0.583
	VGG-19	0.715	0.461	0.573	0.579	0.890	0.968	0.570	0.718

The pixel accuracy, Cohen's kappa, IoU score, AUC, Recall, Precision, and F1 Score of MR-CNN with ResNet101 are better compared to other backbone networks. Based on these parameter values, it can be inferred that the MR-CNN model architecture is the largest model used here in terms of the number of trainable parameters. Consequently, this architecture takes more time to train the system, considering all the used images, than other considered model architectures.

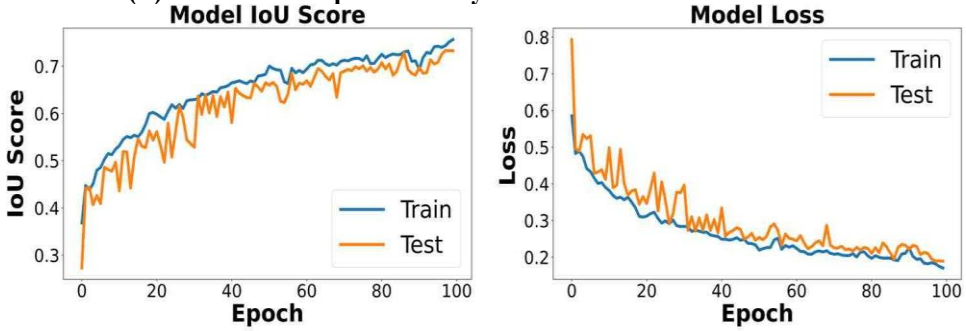
Finally, the Pixel Accuracy with ResNet50, as well as the Cohen's Kappa Score and IoU Score with VGG19 for PSPNet, are better compared to the other two backbone networks. The values of AuC with ResNet50, Recall with ResNet152, Precision, and F1 Score with ResNet50 are the best. It can be inferred that PSPNet performs well with VGG-19, ResNet-50, and ResNet-152 as its backbone networks, respectively.



(a) Score for Prediction by FPN with ResNet50 backbone

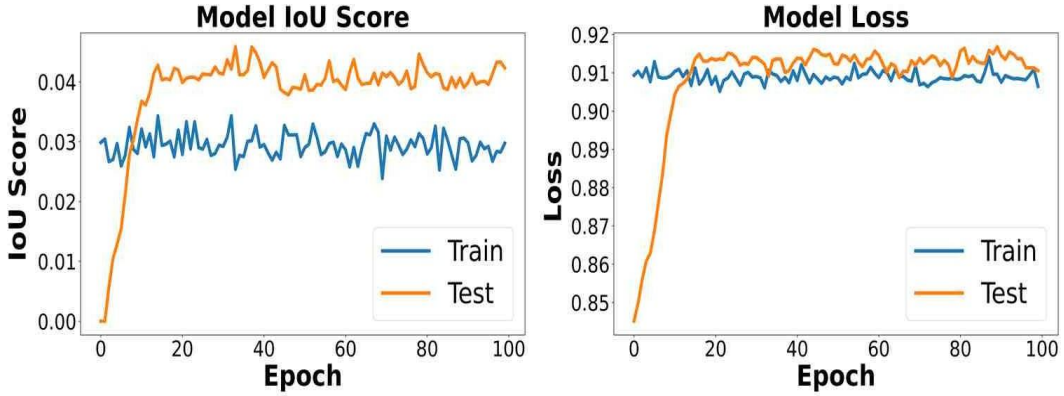


(b) Score for prediction by FPN with ResNet152 backbone

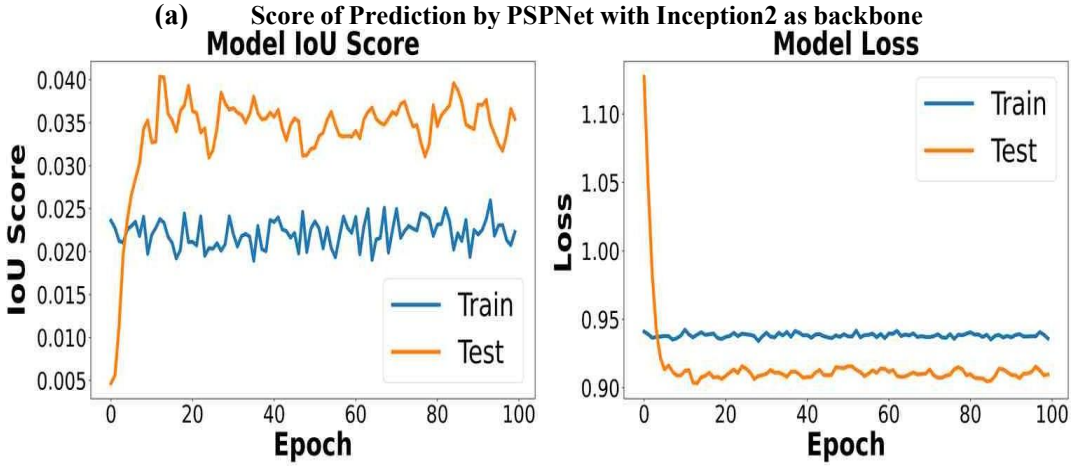


(c) Score for Prediction by FPN with VGG19 backbone

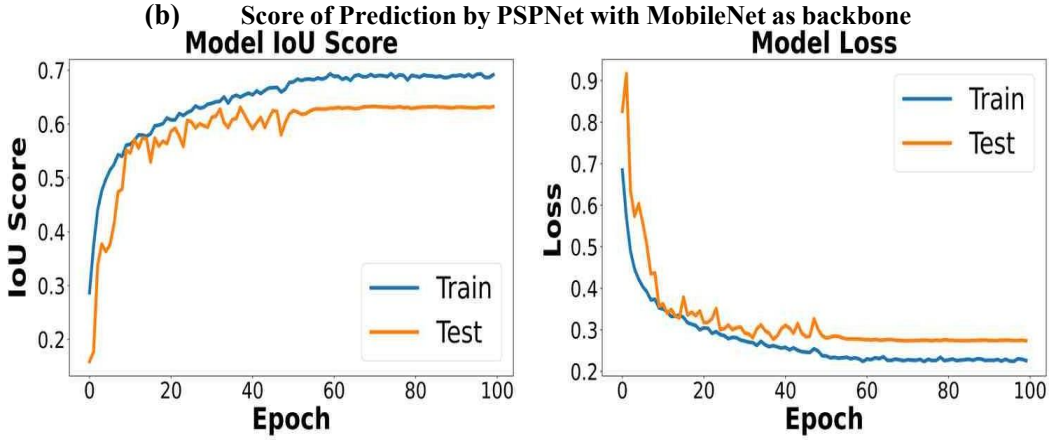
Figure 3: Prediction Scores by FPN



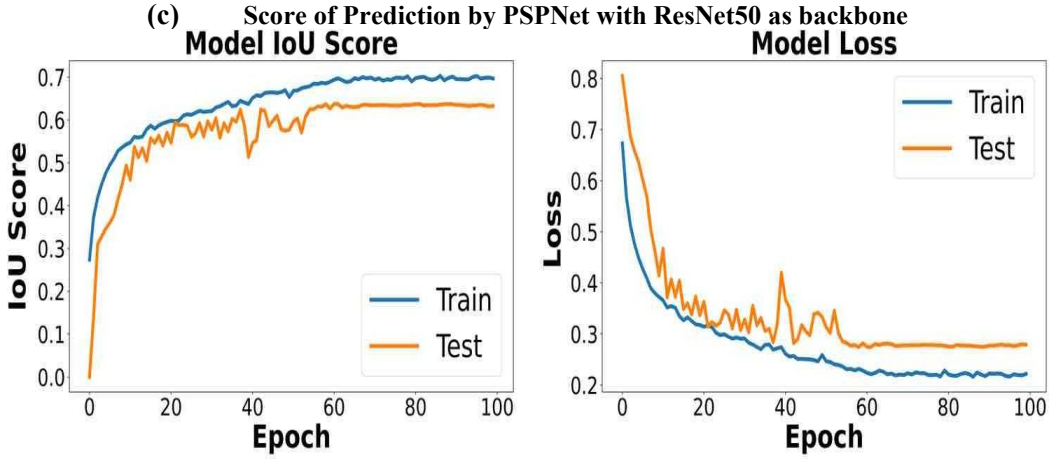
(a) Score of Prediction by PSPNet with Inception2 as backbone



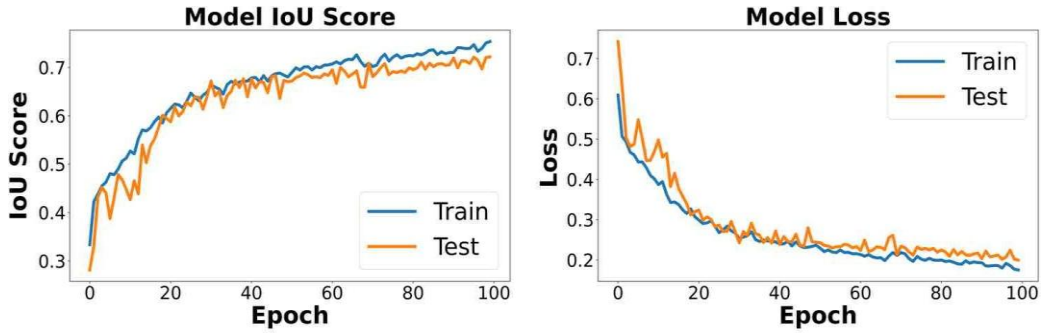
(b) Score of Prediction by PSPNet with MobileNet as backbone



(c) Score of Prediction by PSPNet with ResNet50 as backbone



(d) Score of Prediction by PSPNet with ResNet152 as backbone



(e) Score of Prediction by PSPNet with VCG19 as backbone

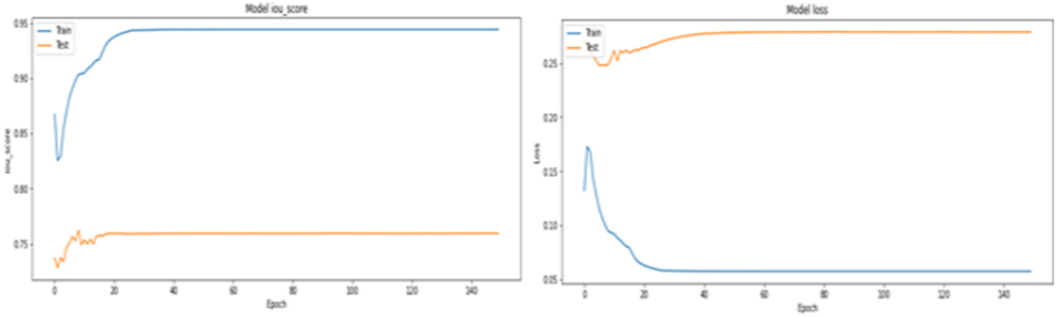


Figure 5: Score of Prediction by M-RCNN

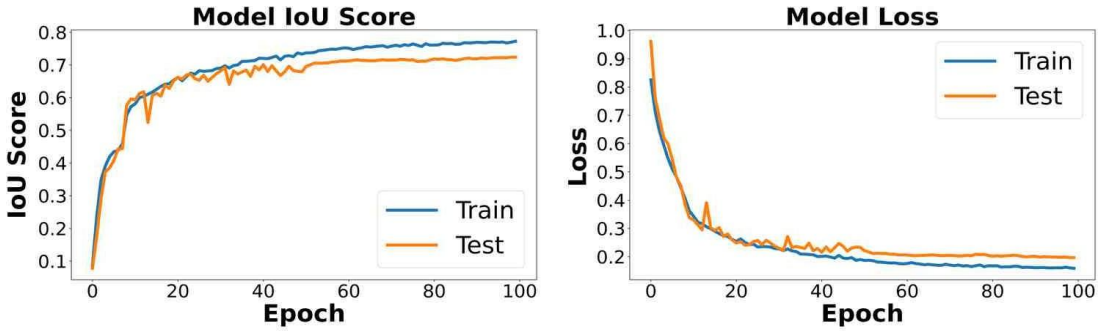
B. Convergence Analysis

The prediction masks obtained using our proposed technique are shown in Figures 3, 4, 5, and 6. It can be seen from the Figures that the top three performing model architectures are FPN with ResNet-152, M-RCNN with ResNet-101, and PSPNet with VGG-19. This is because, as we have seen, the values of the considered parameters for FPN are highest with ResNet152, for M-RCNN are better with ResNet101,and for PSPNet are also best with VGG19. On the other hand, the pixel accuracy, Cohen's kappa, IoU score, AUC, Recall, Precision, and F1 Score of MR-CNN with ResNet101 are better compared to other backbone networks. Finally, we present the convergence analysis of our proposed technique. We show the convergence of the model architectures against each considered backbone network. The CLAHE, Gaussian blur, and various types of augmentation, including translation, Rotation, and Flipping, have been applied. FPN performs prediction with ResNet50, ResNet152, and VGG19 as backbones. FPN with ResNet152 yields the highest score compared to all other models. Similarly, Table 4 shows that mAP, AUC, and F1 Scores for FPN with ResNet152 and VCG19 are the highest, respectively. Therefore, it may be concluded that the FPN gives the best accuracy among all models proposed in this paper. The high F1 score and AuC scores for the top three models confirm that the FPN architecture performs best among all other architectures. It achieves pixel accuracy above 90% for three backbones: ResNet152, VGG-19, and ResNet50. **MRCNN has two backbones due to the computational complexity oflocal computers**

The Pixel Accuracy of 91% for LinkNet with ResNet152 is the highest. The values of Recall andPrecision are highest for LinkNet with ResNet50 and ResNet152, respectively.

The Pixel Accuracy with ResNet50 and Cohen's Kappa Score and IoU Score with VGG19 forPSP-Net are better compared to the other two backbone networks. The values of AuC with ResNet50,Recall with ResNet152, Precision, and F1 Score with ResNet50 are the best.

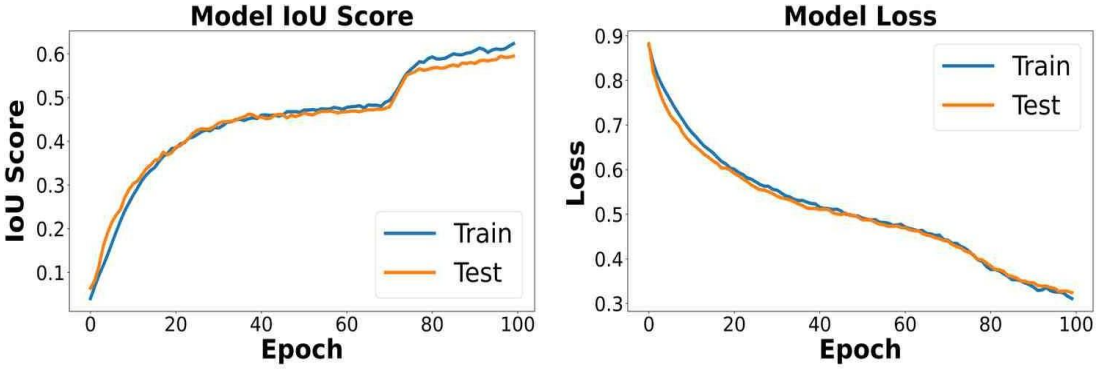
It can be inferred that PSPNet performs well with VGG-19, ResNet-50, and ResNet-152 as its backbone networks, respectively. We also present a graph plotting the convergence time for all model architectures against each backbone network in Figure. LinkNet has the lowest precision, so the prediction mask is omitted.



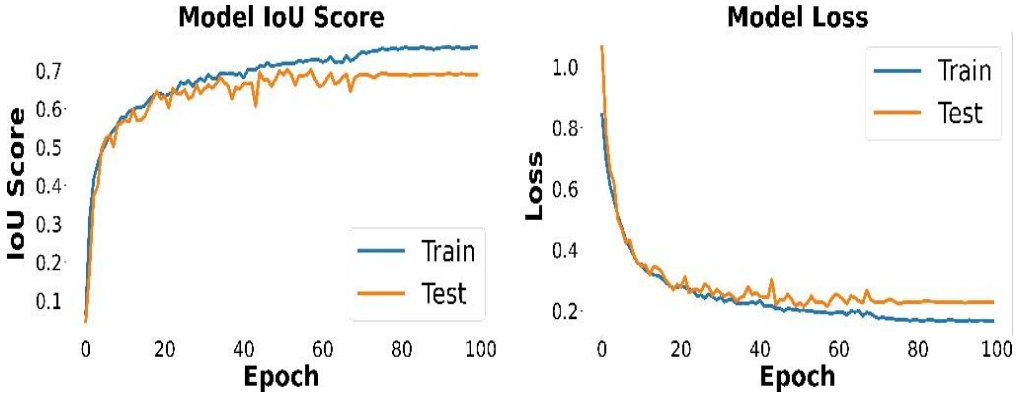
(d) Score of Prediction by LinkNet with Inception3 as backbone

Retrieval Number: 100.1/ijies.E444813050624
DOI: 10.35940/ijies.E4448.11050524
Journal Website: www.ijies.org

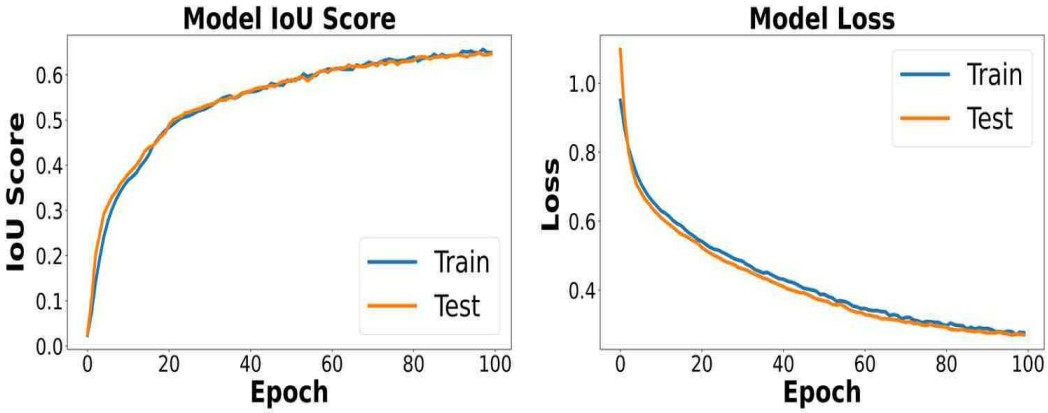
Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)



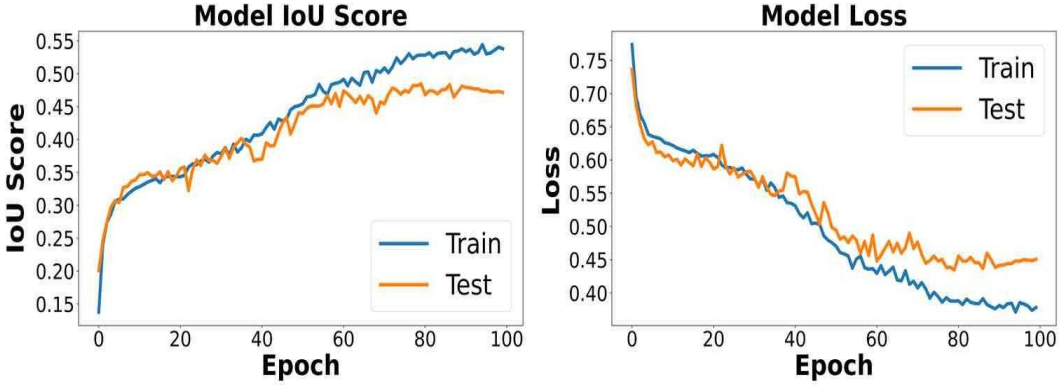
(e) Score of Prediction by LinkNet with MobileNet as backbone



(f) Score of Prediction by LinkNet with ResNet152 as backbone



(g) Score of Prediction by LinkNet with EfficientNet as backbone



(h) Score of Prediction by LinkNet with VGG19 as backbone

C. Case Study

A case study is presented, considering urban images. The accuracy of prediction is shown in Figures 7, 8, and 9 in terms of the mask.

Fig 7: Results of Prediction by FPN with ResNet152 backbone.

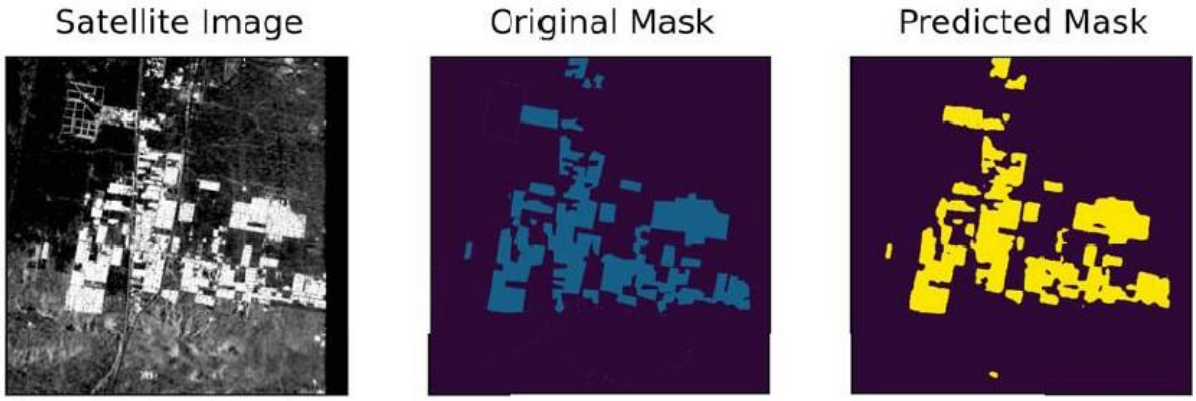


Fig. 8: Results of Prediction by PSPNet

Fig. 9: Results of Prediction by M-RCNN

VI. CONCLUSION

This paper presents an image segmentation and classification technique for urban cover areas using Polarimetric SAR (PolSAR), which is based on Deep Neural Networks (DNNs) such as PSPNet, LinkNet, FPN, and Mask-RCNN. Here, we first applied polarimetric decomposition to airborne Uninhabited Aerial Vehicle Synthetic Aperture (UAVSAR) images of urban areas. The decomposed images were then fed into DNNs for segmentation and classification. Four different experiments are conducted using four distinct databases and models, including PSPNet, LinkNet, FPN, and Mask-RCNN. The results obtained from these experiments are then compared with varying backbone networks, including ResNet-152, ResNet-101, ResNet-50, and VGG-19.In comparison, it is observed that the FPN model with ResNet152 as the backbone network yields the best results on the considered performance metrics, such as mean Average Precision Score (mAP)and pixel accuracy. Specifically, it achieves a pixel accuracy of 90.9% and an mAP score of 0.823, outperforming other Deep Learning models. In the future, the authors would like to explore integrating the proposed technique for change detection and classification of multi-classobjects in the domain of image processing. For a few assets, MrCNN is the best option, and for significant assets, FPN is the most effective ML tool we have used for satellite image segmentation and classification.

Retrieval Number: 100.1/ijies.E444813050624
DOI: 10.35940/ijies.E4448.11050524
Journal Website: www.ijies.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

ACKNOWLEDGEMENT

The authors are grateful to Shivam Aranya, Sayanati Dutta, Raisa Chaterjee, Tripti Kumari, Dr. Tapan Misra, Dr. Arundhati Misra Ray, Dr. Rintu Kumar Gayen, Dr. Sajal Sarkar, and other teachers and students to whom we extend our thanks.

DECLARATION STATEMENT

Funding	No, I did not receive.
Conflicts of Interest	No conflicts of interest to the best of our knowledge.
Ethical Approval and Consent to Participate	No, the article does not require ethical approval or consent to participate, as it presents evidence that is not subject to interpretation.
Availability of Data and Materials	Not relevant.
Authors Contributions	All authors have equal participation in this article.

REFERENCES

1 Z. Niu, G. Hua, X. Gao, *et al.*, "Context aware topic model for scene recognition," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2743–2750, IEEE (2012).
2 S. A. Taghanaki, K. Abhishek, J. P. Cohen, *et al.*, "Deep semantic segmentation of natural and medical images: a review," *Artificial Intelligence Review* **54**(1), 137–178 (2021). https://doi.org/10.1007/s10462-020-09854-1
3 T. Zhou, Z. Li, and J. Pan, "Multi-feature classification of multi-sensor satellite imagery based on dual-polarimetric Sentinel-1a, Landsat-8 OLI, and Hyperion images for urban land-cover classification," *Sensors* **18**(2), 373 (2018). https://doi.org/10.3390/s18020373
4 S.-W. Chen and C.-S. Tao, "Polsar image classification using polarimetric-feature-driven deep convolutional neural network," *IEEE Geoscience and Remote Sensing Letters* **15**(4), 627–631 (2018). https://doi.org/10.1109/LGRS.2018.2799877
5 Y. Zhang, J. Zhang, X. Zhang, *et al.*, "Land cover classification from polarimetric SAR data based on image segmentation and decision trees," Canadian Journal of Remote Sensing 41(1), 40–50 (2015). https://doi.org/10.1080/07038992.2015.1032901
6 S. De, L. Bruzzone, A. Bhattacharya, *et al.*, "A novel technique based on deep learning and a synthetic target database for classification of urban areas in polsar data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **11**(1), 154–170 (2017). https://doi.org/10.1109/JSTARS.2017.2752282
7 S. De, L. Bruzzone, A. Bhattacharya, *et al.*, "A novel technique based on deep learning and a synthetic target database for classification of urban areas in polsar data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **11**(1), 154–170 (2018). https://doi.org/10.1109/JSTARS.2017.2752282
8 Z. Cui, Q. Li, Z. Cao, *et al.*, "Dense attention pyramid networks for multi-scale ship detectionin SAR images," *IEEE Transactions on Geoscience and Remote Sensing* **57**(11), 8983–8997 (2019). https://doi.org/10.1109/TGRS.2019.2923988
9 S. Ren, K. He, R. Girshick, *et al.*, "Faster R-CNN: Towards real-time object detection with region proposal networks," Advances in Neural Information Processing Systems **28**, 91–99 (2015).
10 *S. P. Mohanty, J. Czakon, K. A. Kaczmarek, et al., "Deep learning for understanding satellite imagery:* An experimental survey," *Frontiers in Artificial Intelligence* **3**, 85 (2020). https://doi.org/10.3389/frai.2020.534696
11 X. Wang, L. Zhang, B. Zou, *et al.*, "Polarimetric sar image classification based on kernel sparse representation," in *Compressive Sensing VII: From Diverse Modalities to Big Data Analytics*, **10658**, 106580L, International Society for Optics and Photonics (2018).
12 A. Femin and K. Biju, "Accurate detection of buildings from satellite images using CNN," in*2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, 1–5, IEEE (2020). https://doi.org/10.1109/ICECCE49384.2020.9179232
13 X. Wang, Z. Cao, Z. Cui, *et al.*, "Polsar image classification based on deep polarimetric feature and contextual information," *Journal of Applied Remote Sensing* **13**, 1 (2019). https://doi.org/10.1117/1.JRS.13.034529
14 L. Ding, K. Zheng, D. Lin, *et al.*, "Mp-resnet: Multipath residual network for the semantic segmentation of high-resolution polar images," *IEEE Geoscience and Remote Sensing Letters*, 1–5 (2021). https://doi.org/10.1109/LGRS.2021.3079925
15 L. Zhao and E. Chen, "Segmentation and classification of polar data using spectral graph partitioning," in MIPPR 2013: Remote Sensing Image Processing, Geographic Information Systems, and Other Applications, 8921, 89210E, International Society for Optics and Photonics (2013). https://doi.org/10.1117/12.2031128
16 A. Ouahabi and A. Taleb-Ahmed, "Deep learning for real-time semantic segmentation: Application in ultrasound imaging," *Pattern Recognition Letters* **144**, 27–34 (2021). https://doi.org/10.1016/j.patrec.2021.01.010
17 Y. Chen, X. He, J. Wang, et al., "The influence of polarimetric parameters and an object-based approach on land cover classification in coastal wetlands," *Remote Sensing* **6**(12), 12575– 12592 (2014). https://doi.org/10.3390/rs61212575
18 K. He, X. Zhang, S. Ren, *et al.*, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778 (2016).
19 S. Xie, R. Girshick, P. Dolla´r, et al., "Aggregated residual transformations for deep neural networks," *arXiv preprint arXiv:1611.05431* (2016). https://doi.org/10.1109/CVPR.2017.634
20 K. Simonyan and A. Zisserman, "Intense convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556* (2014).
21 P. Burlina, "Mrcnn: A stateful fast r-cnn," in *2016 23rd International Conference on Pattern Recognition (ICPR)*, 3518–3523 (2016). https://doi.org/10.1109/ICPR.2016.7900179
22 *T.-Y. Lin, P. Dolla´r, R. Girshick, et al., "Feature pyramid networks for object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition*, 2117–2125(2017).
23 A. Chaurasia and E. Culurciello, "Linknet: Exploiting encoder representations for efficient semantic segmentation," in *2017 IEEE Visual Communications and Image Processing (VCIP)*, 1–4, IEEE (2017). https://doi.org/10.1109/VCIP.2017.8305148
24 H. Zhao, J. Shi, X. Qi, *et al.*, "Pyramid scene parsing network," in *Proceedings of the IEEEconference on computer vision and pattern recognition*, 2881–2890 (2017). https://doi.org/10.1109/CVPR.2017.660
25 S. Cloude and E. Pottier, "A review of target decomposition theorems in radar polarimetry," *IEEE Transactions on Geoscience and Remote Sensing* **34**(2), 498–518 (1996). https://doi.org/10.1109/36.485127
26 E. Pottier and J.-S. Lee, "Application of the h/a/alpha polarimetric decomposition theorem for unsupervised classification of fully polarimetric SAR data based on the Wishart distribution," in *SAR workshop: CEOS Committee on Earth Observation Satellites*, **450**, 335 (2000).
27 M. Neumann, L. Ferro-Famil, and E. Pottier, "A general model-based polarimetric decomposition scheme for vegetated areas," in Proceedings of the 4th International Workshop on Science and Applications of SAR Polarimetry and Polarimetric Interferometry (ESRIN), Frascati, Italy, 26–30, Citeseer (2009).
28 A. Freeman, "Fitting a two-component scattering model to polarimetric SAR data from forests," IEEE Transactions on *Geoscience and Remote Sensing* **45**(8), 2583–2592 (2007). https://doi.org/10.1109/TGRS.2007.897929
29 A. Freeman and S. Durden, "A three-component scattering model for polarimetric SAR data," *IEEE Transactions on Geoscience and Remote Sensing* **36**(3), 963–973 (1998). https://doi.org/10.1109/36.673687
30 J. R. Huynen, "Stokes matrix parameters and their interpretation in terms of physical target properties," in *Polarimetry: Radar, infrared, visible, ultraviolet, and X-ray*, **1317**, 195–207,International Society for Optics and Photonics (1990). https://doi.org/10.1117/12.22083
31 A. Bhattacharya, A. Muhuri, S. De, *et al.*, "Modifying the Yamaguchi four-component decomposition scattering powers using a stochastic distance," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **8**(7), 3497–3506 (2015). https://doi.org/10.1109/JSTARS.2015.2420683
32 G. Singh and Y. Yamaguchi, "Model-based six-component scattering matrix power decomposition," *IEEE Transactions on Geoscience and Remote Sensing* **56**(10), 5687–5704 (2018).

https://doi.org/10.1109/TGRS.2018.2824322
33 R. Barnes, "Roll-invariant decompositions for the polarisation covariance matrix," in Proceedings *of the Polarimetry Technology Workshop, Redstone Arsenal, AL, USA*, **1618** (1988).
34 W. A. Holm and R. M. Barnes, "On radar polarization mixed target state decomposition techniques," in *Proceedings of the 1988 IEEE National Radar Conference*, 249–254, IEEE (1988).
35 M. Arii, J. J. van Zyl, and Y. Kim, "Adaptive model-based decomposition of polarimetric SAR covariance matrices," *IEEE Transactions on Geoscience and Remote Sensing* **49**(3), 1104– 1113 (2011). https://doi.org/10.1109/TGRS.2010.2076285
36 W. An, Y. Cui, and J. Yang, "Three-component model-based decomposition for polarimetric SAR data," *IEEE Transactions on Geoscience and Remote Sensing* **48**(6), 2732–2739 (2010). https://doi.org/10.1109/TGRS.2010.2041242
37 W. An, C. Xie, X. Yuan, *et al.*, "Four-component decomposition of polarimetric SAR images with deorientation," *IEEE Geoscience and Remote Sensing Letters* **8**(6), 1090–1094 (2011). https://doi.org/10.1109/LGRS.2011.2157078
38 *Y. Yamaguchi, G. Singh, C. Yi, et al., "Comparison of model-based four-component scattering power decompositions," in Conference Proceedings of 2013 Asia-Pacific Conference on Synthetic Aperture Radar (APSAR), 92–95, IEEE (2013).*
39 Y. Yamaguchi, T. Moriyama, M. Ishido, *et al.*, "Four-component scattering model for polarimetric SAR image decomposition," *IEEE Transactions on Geoscience and Remote Sensing* **43**(8), 1699–1706 (2005). https://doi.org/10.1109/TGRS.2005.852084
40 E. Pottier, F. Sarti, M. Fitrzyk, *et al.*, "Polsarpro-biomass edition: The new ESA polarimetric SAR data processing and educational toolbox for the future ESA & third party fully polarimetric SAR missions," in *ESA Living Planet Symposium 2019*, (2019).

AUTHORS PROFILE

Tamesh Halder, Department of Mining Engineering Indian Institute of Technology Kharagpur, India Unmanned Aerial Vehicles, Synthetic Aperture Radar, Ad Hoc Networks, Agroforestry, Digital Elevation Model, Extreme Learning Machine, Flying Ad Hoc Networks, Global Positioning System, Path Loss, Point Cloud, Point Cloud Model, Polarimetric Synthetic Aperture Radar, Receiver Operating Characteristic Curve,3D Mesh, Acid Mine Drainage, Acquisition Unit, Adaptive Algorithm, Adaptive Filter, Additive Noise, Aerial Images, Aerodynamic, Agisoft Photo Scan, Area Under Curve, Artifact Removal, Atmospheric Oxygen Tamesh Halder received the B.Tech. He received a degree in Electronics and Instrumentation Engineering from the University of Kalyani in 2010 and an M.S. degree in Electronics and Electrical Communications Engineering from the Indian Institute of Technology, Kharagpur, in 2013. He is currently pursuing a Ph.D. degree in the Department of Mining Engineering. His research experience includes remote sensing, focal source localization, compressive sensing, beamformers, receiver operating characteristics, EEG, MEG, and machine learning.

Soumyadip Sarkar, Activities: ex-ML intern @IIT Kharagpur | Junior ML researcher at IEDC Lab | Kaggle Kernel Expert |Deep Learning | Computer Vision | NLP | Love Python and C++. Summary: Hello, it's Soumyadip Sarkar. I graduated in 2022 with a degree in Electrical Engineering, IEM Kolkata, India. I have experience in Computer Vision, NLP, Machine Learning, and Deep Learning. I am familiar with backend development for ML/DL models using Flask, Streamlit, and Docker. I love participating in Kaggle competitions and different Hackathons. I also enjoy making open-source contributions and writing blogs. Performed different qualitative and quantitative experiments on the Mask RCNN model for image segmentation of PolSAR data.

Farhan Hai Khan, ML Head at Applex. in | Jr. ML Researcher at IEM IEDC | Kaggle Expert | ex-ML Intern IIT KGP | ex-Data Science Intern BrickView Studios | WoC 20 and GSSoC 21 Mentor | AI Team @DSC-IEM | Core IET IEM | IEM EE Hello, this is Farhan Hai Khan, my bachelor's degree in Electrical Engineering from the Institute of Engineering and Management, Kolkata in 2022. I'm a Machine Learning enthusiast, also skilled in algorithm design and Project Work. Developing Deep Learning skills this year. Work done during the internship period: Polarimetric Orientation Angle Estimation using various Stochastic Distancing Techniques.Hellinger, Renyi of Order Beta, Wasserstein, Bhattacharya Distances. Remote Sensing-based GeoData.

Shobhit Kumar, Activities: ex-Summer intern at Tata Steel (Tech Team), Ex-Research intern @iit kharagpur || CP(3 ⭐@codechef)||Data Science from IIT Madras || ML enthusiast|| DevOps Hello, it's Shobhit Kumar, studied at IEM Kolkata, India since Nov 2020 pursuing BTech at computer science. I have experience in Computer Vision, NLP, Machine Learning, and Deep Learning.

Dipjyoti Paul, Affiliation Department of Computer Science, University of Crete, Rethymno, Greece, Publication Topics, Generative Adversarial Networks, Convolutional Layers, Fréchet Inception Distance, Generative Adversarial Networks Training, Nash Equilibrium, Neural Network, Adult Male, Baseline Training, Batch Normalization, Convergence Rate, Covariance Matrix, Deep Neural Network, F-statistic, Gaussian Mixture Model, Generative Adversarial Networks Loss, Generator Training, Hellinger Distance, High-quality, Higher-order Statistics, Hourly Data, Hyperparameter Values, Image Generation, Inception Distance, Kullback-Leibler, Learning Rate Dipjyoti Paul received the B.Tech. Degree in electronics and communication engineering from the St. Thomas' College of Engineering and Technology under the West Bengal University of Technology, Kolkata, India, in 2013, and the M.Sc. degree in electronics and electrical communication engineering from the Indian Institute of Technology Kharagpur (IIT Kharagpur), Kharagpur, India, in 2017. He is currently pursuing a Ph.D. degree in speech signal processing at the Department of Computer Science, University of Crete, Rethymno, Greece. His research interests encompass diverse areas, including machine learning, deep learning, signal processing, speech processing, computer vision, and optimisation theory. He also has a broad interest in probabilistic machine learning methods and generative models. Paul is also a member of ISCA.

Debashish Chakravarty, Affiliation Mining Engineering Indian Institute of Technology Kharagpur Kharagpur, India Publication Topics Unmanned Aerial Vehicles, Global Positioning System, Synthetic Aperture Radar, Ad Hoc Networks, Agroforestry, Convolutional Neural Network, Digital Elevation Model, EEG Data, Extreme Learning Machine, Flying Ad Hoc Networks, Land Use, Land Use/land Cover, Machine Learning, Mining Regions, Normalized Difference Vegetation Index, Path Loss, Performance Metrics, Point Cloud, Point Cloud Model, Polarimetric Synthetic Aperture Radar, Receiver Operating Characteristic Curve, Water Bodies,3D Mesh, Acid Mine Drainage, Acquisition Unit Debashish Chakravarty received the B.Tech. and M.Tech. Degrees in mining engineering, a Ph.D. degree from the Indian Institute of Technology Kharagpur, Kharagpur, in 1998, and a postdoctoral degree in applied mathematics and computer science from BAT, Germany, in 2003. He is an associate professor in the Department of Mining Engineering at the Indian Institute of Technology Kharagpur. He is the Professor-in-Charge of the Autonomous Ground Vehicle Students' Research Group. He is also affiliated with the Advanced Technology Development Centre at the Indian Institute of Technology Kharagpur. His areas of research are remote sensing, rock mechanics, and numerical modelling